

Supplementary Information to

Resolving Conflict in Eutherian Mammal Phylogeny Using Phylogenomics and the Multispecies Coalescent Model

Sen Song^{a,1}, Liang Liu^{b,1}, Scott V. Edwards^c, Shaoyuan Wu^{d,b,2}

^aDepartment of Biomedical Engineering, School of Medicine, Tsinghua University,
Beijing 100084, China.

^bDepartment of Statistics and Institute of Bioinformatics, University of Georgia, Athens,
GA 30606, USA

^cDepartment of Organismic and Evolutionary Biology and Museum of Comparative
Zoology, Harvard University, Cambridge, MA 02138, USA

^dInstitute of Paleontology, Shenyang Normal University, Shenyang 110034, China

¹These authors contributed equally to this work

²To whom correspondence should be addressed: Email: shaoyuan5@gmail.com

SI Materials and Methods

Taxon and gene sampling. We examined 33 eutherian taxa, representing members from 16 eutherian orders. These taxa were selected because they have whole genome sequences available in ENSEMBL. The outgroups include two marsupials, one monotreme and one non-mammalian taxon chicken. The taxa included in this study and their classifications are provided in Table S1.

We conducted an exhaustive search for CDS in the OrthoMam database (1), using the criterion that the orthologous genes should be available for all the mammalian taxa sampled. We found 796 genes that meet this criterion, and the CDS and ML gene trees for those genes were downloaded from the OrthoMam database (June 2011). We further reduced the dataset to 706 genes by requiring those to also have orthologous CDS for the outgroup chicken. From this gene set, we excluded genes that contained long-branches, strong violations of the molecular clock, or likely paralogs. We measured the distance from the root to each of the tip nodes for the 706 gene trees by examining the ML gene trees downloaded from OrthoMam using a custom script written in python using the PyCogent (2) and Biopython packages (www.biopython.org). We excluded trees in which the maximum distance to a given tip is three times or longer than the average branch length. As a result, 447 genes with a total length of 1,385,220 bp were retained for this study. The genes sampled and their distribution across human chromosomes are provided in Table S2.

Since OrthoMam do not contain chicken sequences, we downloaded the chicken orthologous sequences directly from the ENSEMBL database (release 62) with python scripts using the PyCogent (2) and Biopython packages (www.biopython.org).

Sequence alignment: We used MUSCLE 3.7 (3, 4) to perform multiple alignments for each of the genes selected across all 37 taxa. Custom python scripts were written to automate this process.

Model Selection: The best-fit substitution model for each of the 447 genes was selected by the Akaike Information Criterion. The log-likelihoods of the substitution models for 447 genes were obtained in RAxML (5) and then used to calculate $AIC = -2\log\text{likelihood} + 2P$, where P is the number of parameters in the model. It was suggested by the Akaike Information Criterion that GTR+ Γ is the best-fit model for 364 genes, while TIM+ Γ is the best-fit model for the remaining 83 genes. Additionally, the second best-fit model for the 83 genes is GTR+ Γ and the difference of the AIC between the two models, GTR+ Γ and TIM+ Γ , is less than 3. We constructed gene trees for each of those 83 genes using both GTR+ Γ and TIM+ Γ models, and the gene trees based on both models are identical in topologies. Thus, GTR+ Γ is selected as the substitution model used in the concatenation and coalescent methods for reconstructing the phylogeny of eutherian mammals.

Test of multispecies coalescent model

Gene tree distances: We evaluated how well the multispecies coalescent model can explain the gene tree variation observed in the data set of 447 genes. We first built a MP-EST tree with branch lengths in coalescent units from the data set and then simulated 10,000 gene trees from the MP-EST species tree under the coalescent model. The 10,000 gene trees served as the gene trees expected from the coalescent model.

We calculated the Robinson-Foulds (6) distance between the 447 simulated gene trees and the MP-EST tree. The distances are used to measure the expected “gene tree variation”. We then calculated the distance between the observed gene trees (those estimated from the 447 genes) and the MP-EST tree. The distances are used to measure the observed “gene tree variations”. We calculated the ratio between the median of the expected distances and the median of the observed distances, where the median distance is used to measure the overall gene tree variation. The ratio is equal to $20/26=0.77$, indicating that the coalescent model can account for 77% of the gene tree variation observed in the real data set of 447 gene trees (Fig. 3a).

Consensus tree: The expected gene trees were summarized by a majority-rule consensus tree in PHYLIP (7, 8), which is the consensus tree expected from the coalescent model. We also built a consensus tree from the observed gene trees estimated from each of the 447 genes in the real data set. The expected consensus tree is topologically consistent with the observed consensus tree (Fig. S4). Moreover, the percentages at the internal nodes of the expected consensus tree are similar to those in the observed consensus tree (Fig. S4). The percentages in the expected consensus tree represent the gene tree variation expected from the coalescent model.

Frequency of gene tree triplets: We also counted the frequency of gene trees from triplets of taxa as a test of the multispecies coalescent model. This test is outlined in Ané (9), and compares the frequency of the two minority gene trees in the triplet with that expected from the multispecies coalescent. Coalescent theory does not predict the absolute frequency of minority triplets or their relative frequency compared to the majority triplet – these numbers depend on the internal branch length of the triplet in the species tree as measured in coalescent units. But the theory does predict that these minority gene trees should be equally frequent.

Because the frequency of rooted gene tree triplets are not independent of one another when each tip corresponds to a clade of species rather than a single species, we can summarize the frequencies by counting for each gene the number of gene tree triplets supporting ((speciesA, speciesB), speciesC); ((speciesA, speciesC), speciesB); or (speciesC, speciesB), speciesA), such that if the majority of triplets for the clades support ((speciesA, speciesB), speciesC), we count 1 for that triplet for that gene.

The test assumes that gene trees are given without error. This may lead to underestimation of the variance of the test statistic. Still, if we consider the variation in estimating gene trees, the frequencies are likely within the range of their expectation. It is ultimately difficult to measure the effect of the estimation error in gene trees on the test statistic. The test also implicitly assumes that the most probable gene tree triplet is

congruent with the corresponding species tree triplet. This assumption will usually hold provided that the subtree under investigation is not in the anomaly zone (9). Regardless, whether or not this assumption is met, MP-EST will still work well because although it uses the most frequent gene tree triplet (the other two minority frequencies are much less relevant to the MP-EST method) to estimate the species tree triplet, it has also been shown to perform well even in the anomaly zone (10).

Test for systematic effect of recombination: We examined the effect of recombination in our study, particularly among longer loci. The loci we sampled are long for a phylogenetic study (mean = 3126). We have only sampled single sequences per species, thus a traditional test of recombination, which requires multiple sequences per species, is not possible with these data. Still, recombination can be detected in multispecies data if one assumes that recombinant sequences could be generated in ancestral populations, where sequences from two species coalesce and are present in the same population. If recombination was a factor in our sequences, we would expect that the consistency index, a measure of homoplasy (11), would be higher in longer sequences, since these allow for more opportunities for recombination along the length of the sequence. The average consistency index among our loci was 0.413 (± 0.037 , 1 s.d.). We find that there is no correlation between gene length and consistency index ($t = -0.4048$, $df = 445$, $P\text{-value} = 0.686$), suggesting that our long loci incur no more recombination than shorter loci in our data set (Fig. S10).

Test for the relationships between gene length and the average bootstrap value. We estimated gene trees with 100 bootstrap replicates for each of the 447 genes using RAxML, and then built a consensus tree for each of the 447 genes with the method mentioned above. We calculated the average bootstrap value on the consensus tree for each gene. We performed a log transformation of the average bootstrap values and the gene lengths, and conducted a linear regression of average bootstrap values on the gene lengths, as plotted in Fig. S11. We also performed MP-EST analysis for the eutherian mammal tree with the longest and the shortest 50, 100 and 200 gene sets from the 447 gene set, and calculated the average bootstrap values for each of the trees (Table S5).

Estimation of eutherian phylogeny using the data set of Meredith et al. 2011 (12):

The sampling of taxa in Meredith et al. 2011 shares 31 taxa with that of our study. These taxa include 27 eutherian taxa, representing 16 out of 18 eutherian orders. We performed phylogenetic analyses based on those 31 taxa and the genetic data set of Meredith et al. 2011 using both coalescent and concatenation methods. We obtained the genetic data set used in Meredith et al. 2011 from the nexus file that they submitted to treebase (www.treebase.org). We employed two coalescent methods, MP-EST and STAR, and the concatenation method in MrBayes, to this data. The 31 taxa included for these analyses are provided in Table S4.

SI Text

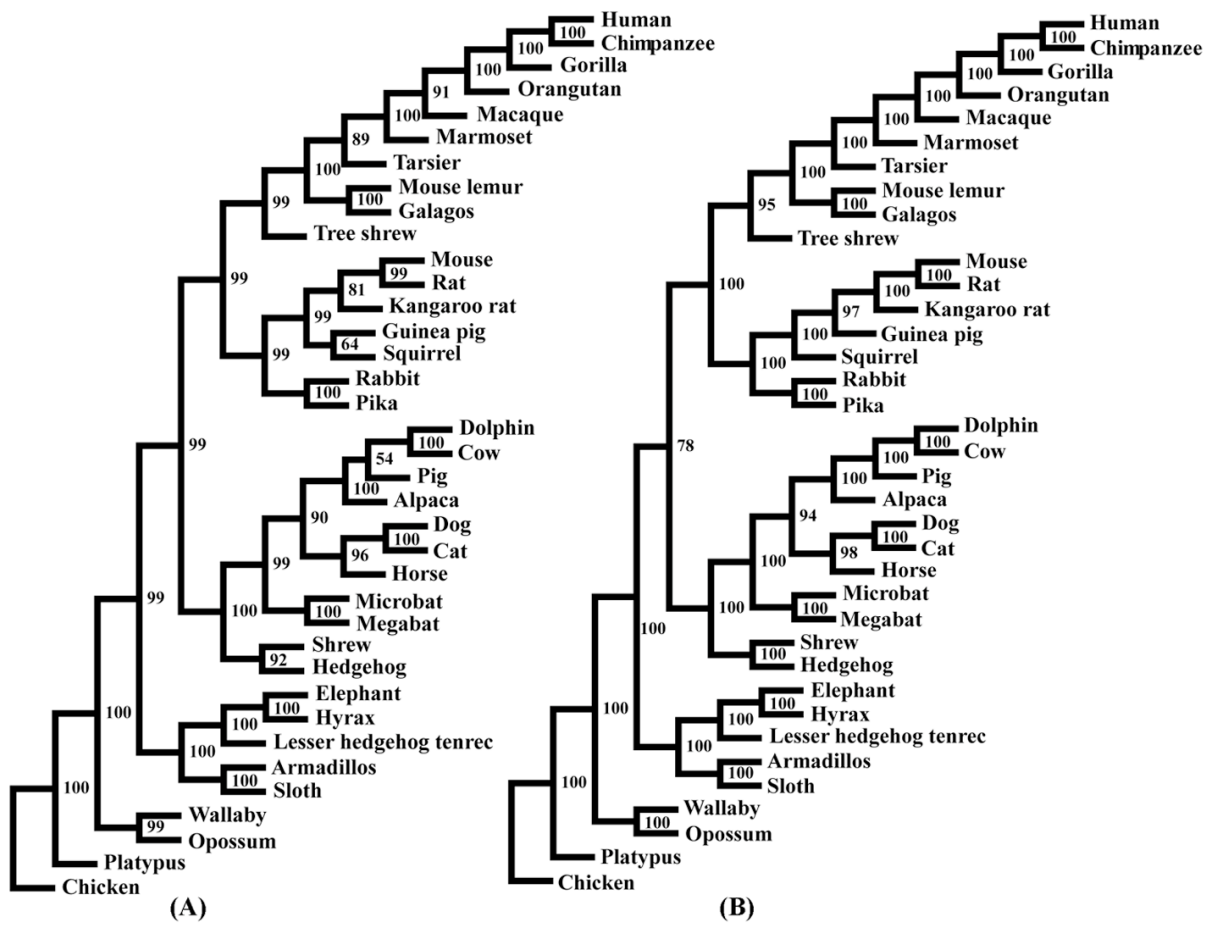
Results of phylogenetic analyses with reduced taxon sampling

Coalescent analyses: Eutherian phylogenies estimated with two reduced data sets are consistent with the result given by the original data set (Figs. S5-S6). When twelve taxa excluded, the relationship among the three rodent species, squirrel, rat and kangaroo rat, remain unresolved in the MPEST tree, as the group of squirrel and kangaroo rat received a low bootstrap support value of 68%.

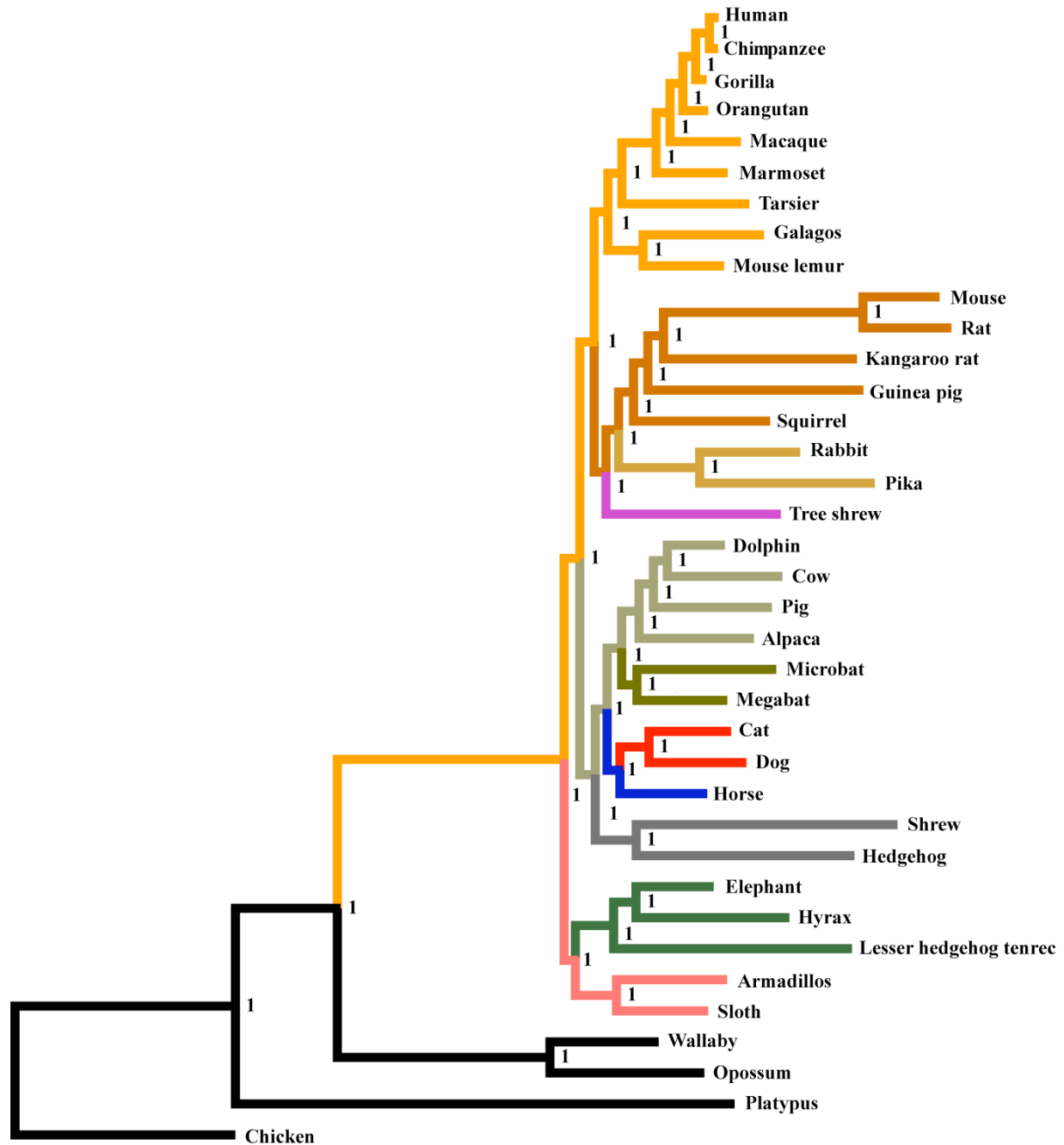
Concatenation analyses: Eutherian phylogenies using reduced taxon sampling are inconsistent with the one from the full taxon set (Figs. S7-S8). Tree shrew is grouped with primates in the tree estimated using Bayesian method with reduced data set (posterior probability (PP)=93% when six taxa excluded and 100% when 12 taxa excluded) (Fig. S7). By contrast, in the tree estimated using the full data set, tree shrew is the sister taxon of Glires (PP=100%) (Fig. S2). For ML analyses, tree shrew is grouped with Glires for the full taxon set (BP=0.95) and the exclusion of six taxa (BP=0.9), whereas tree shrew is grouped with primates when twelve taxa excluded (BP=0.9) (Fig. S3B and Fig. S8). Those conflicting phylogenies received high nodal support values, suggesting that taxon sampling can also be one source of phylogenetic bias for concatenation methods, resulting in misleading phylogeny.

The taxon sampling of our analyses does not include members of two eutherian orders, Dermoptera (colugos) and Pholidota (pangolin), because no whole genome sequences are available for them so far. However, although proof of consistency will come with increased taxon sampling, the robustness of coalescent methods to variable taxon sampling suggests that the lack of these two orders should not have a significant impact on the phylogenetic framework estimated by this study for the interordinal relationships among eutherian mammals.

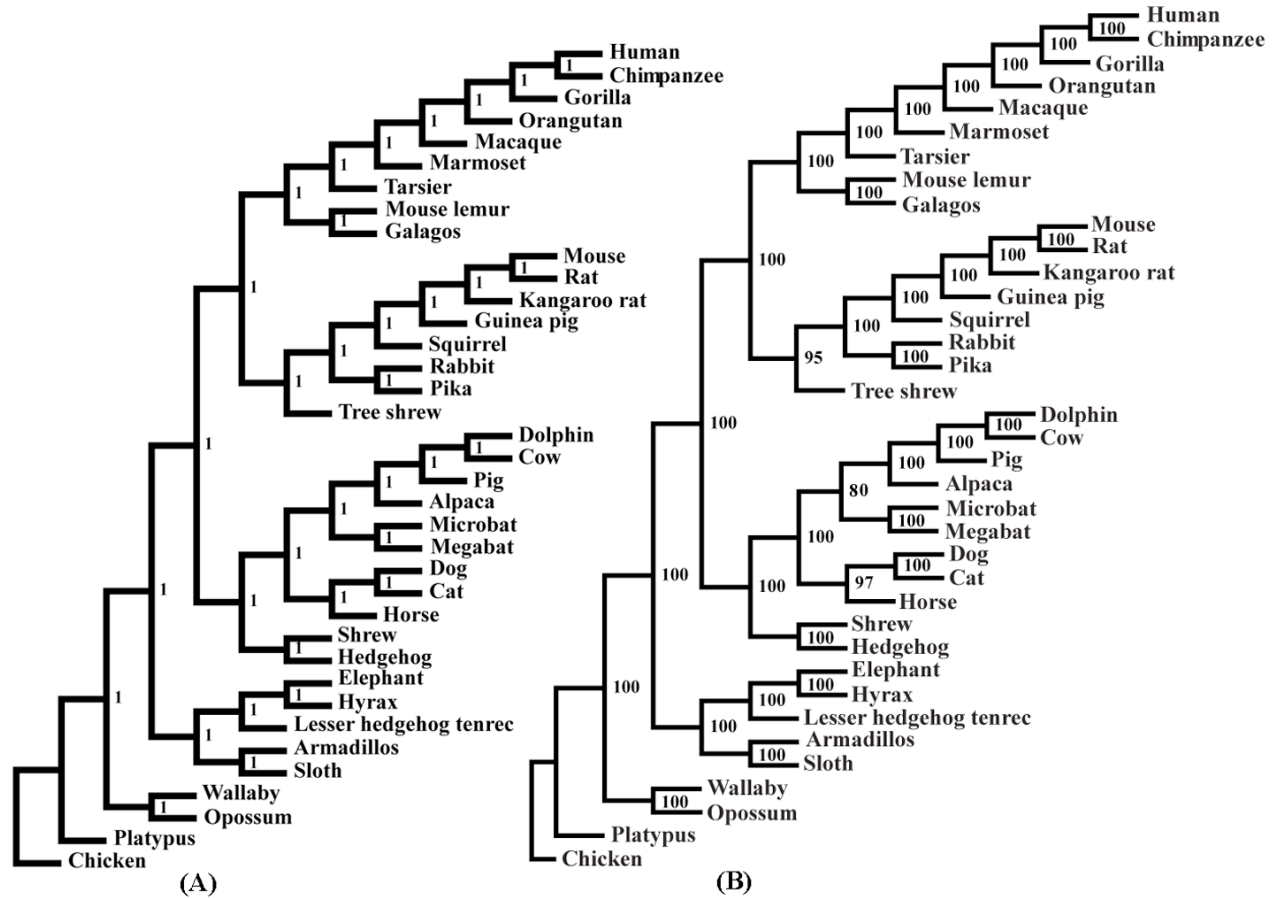
SI Figures



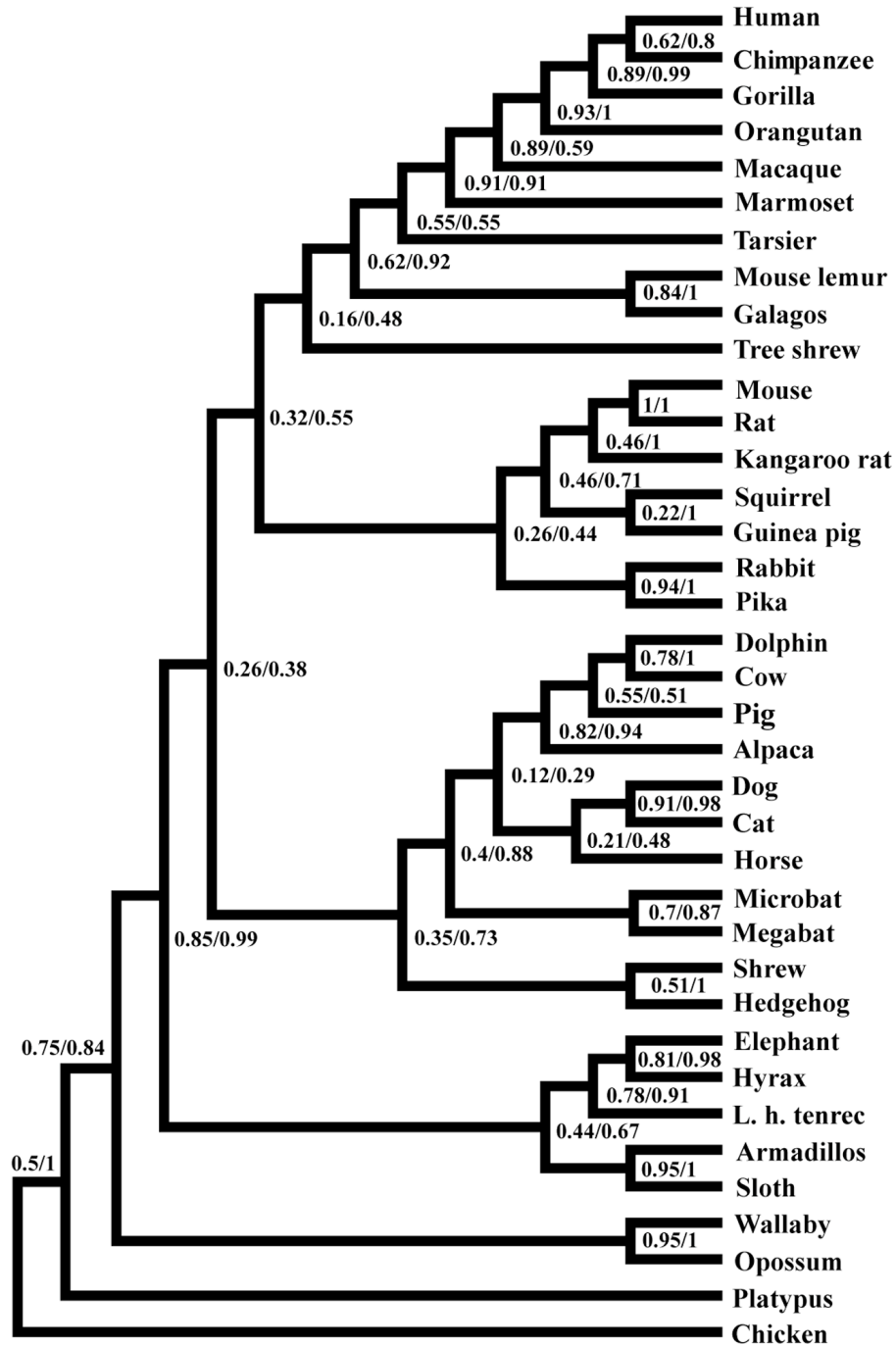
Supplementary Fig. S1. Phylogenetic relationships of eutherian mammals estimated using the coalescent methods MP-EST and STAR for 447 genes. A. MP-EST tree; B. STAR tree. The numbers on the tree indicate the bootstrap support values for each of the nodes.



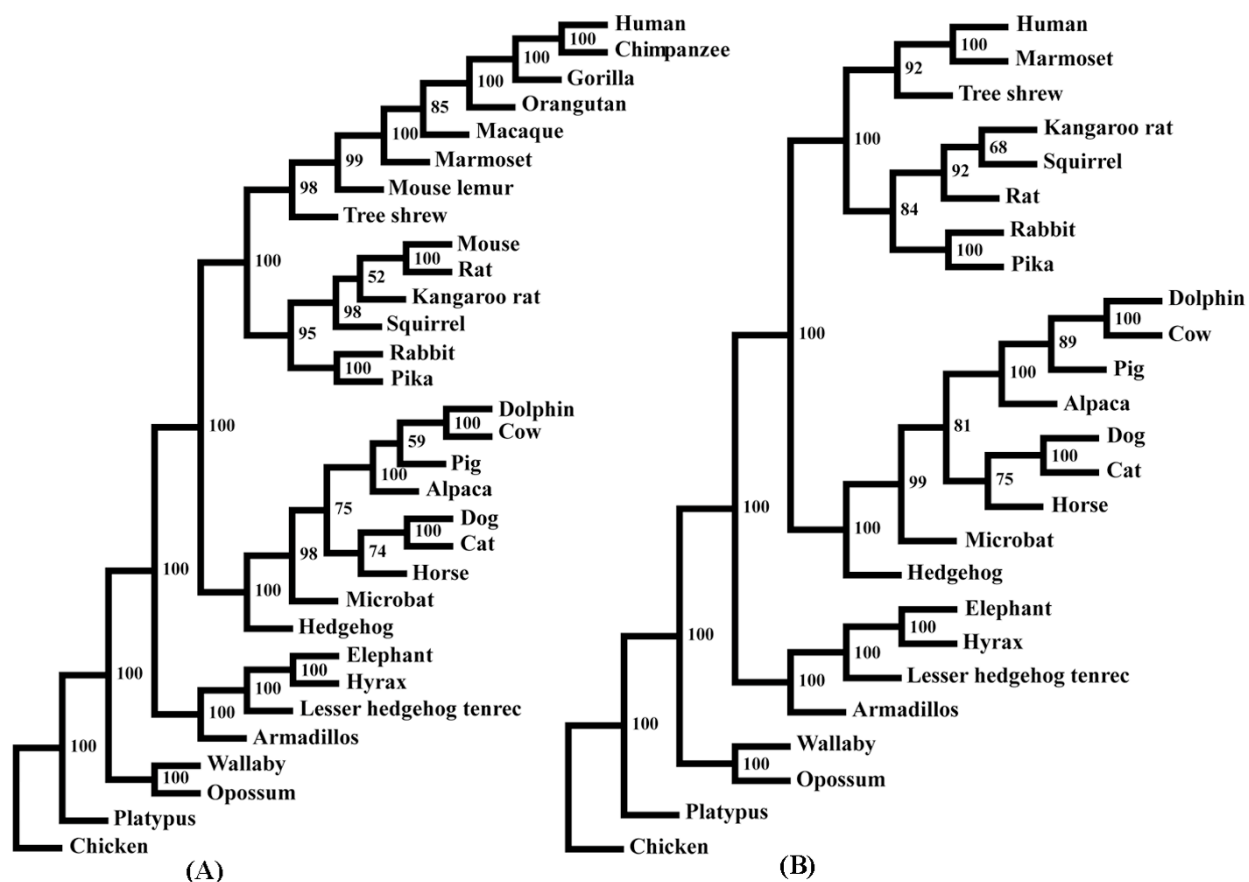
Supplementary Fig. S2. Phylogenetic relationships of eutherian mammals estimated using the concatenation method MrBayes for 447 genes. Numbers on the tree indicate the posterior probability percentages for each of the node. This tree is the same as provided in the inset of Figure 1 of the main text.



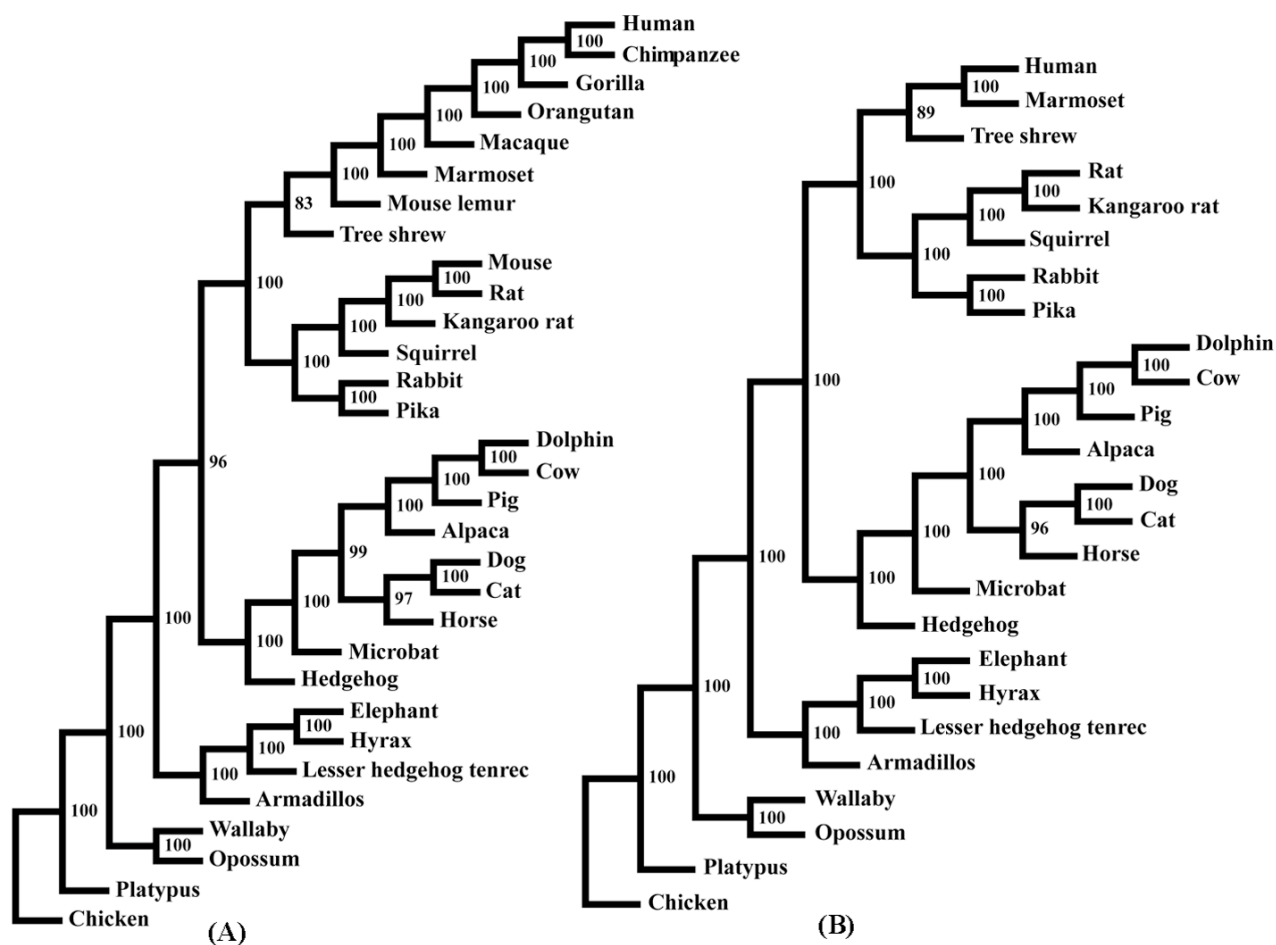
Supplementary Fig. S3. Phylogenetic relationships of eutherian mammals estimated using the concatenation methods MrBayes (Bayesian) and RAXML (maximum likelihood) for 447 genes. A. MrBayes tree; B. RAXML tree. The numbers on the tree indicate the posterior probability or bootstrap values for each of the nodes.



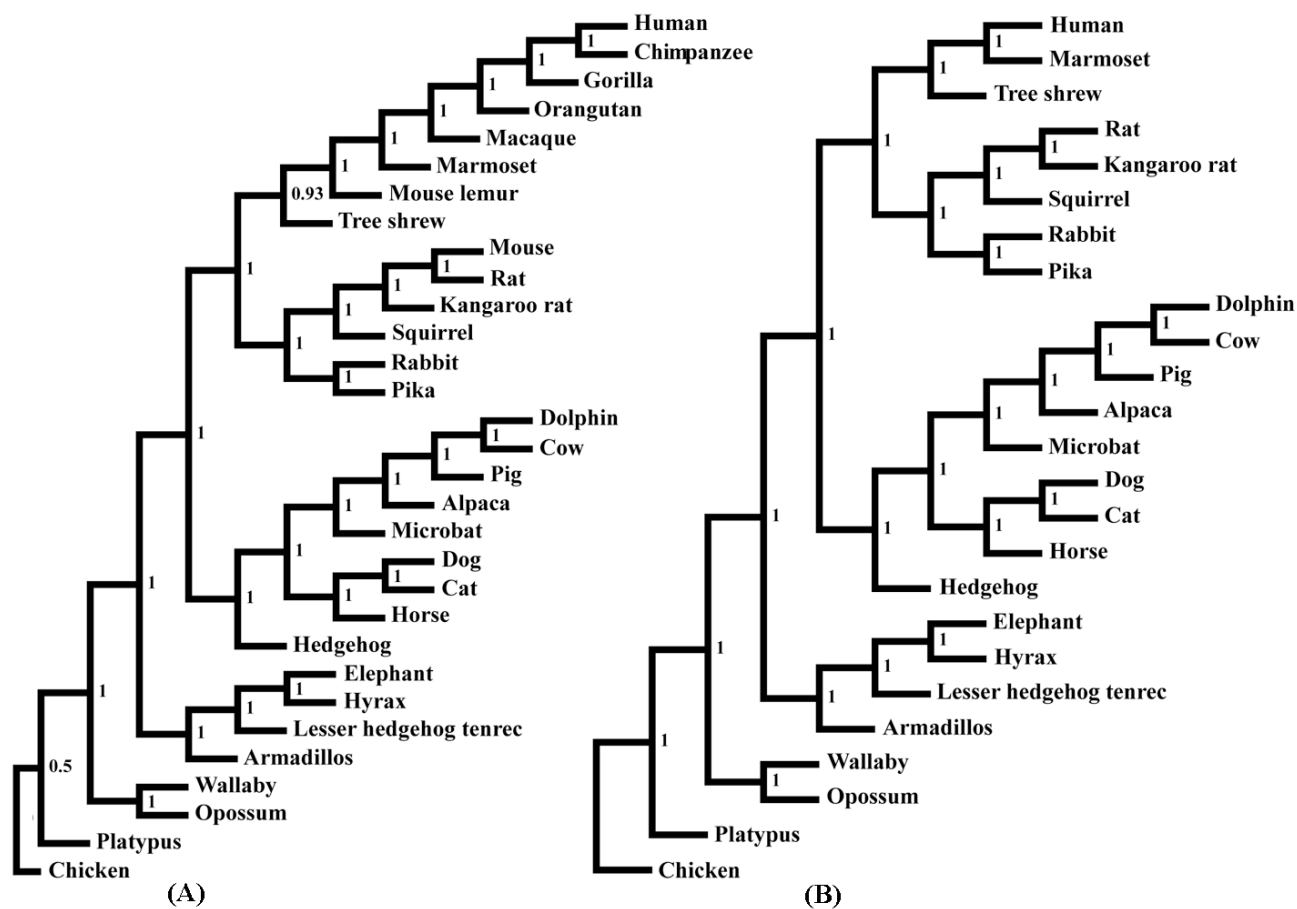
Supplementary Fig. S4. The consensus gene tree indicates the extent of gene tree heterogeneity. The consensus gene tree estimated from the 447 individual gene trees of the real data and the 10,000 individual gene trees simulated from the MP-EST species tree based on the coalescent model. Both consensus gene trees are identical in topology, which is also identical to the MP-EST species tree. The numbers on the tree indicate the fraction of gene trees that supports the placement of the corresponding nodes from the real data and the simulations on the left and right of the slash respectively. Abbreviation: L. h. tenrec, lesser hedgehog tenrec.



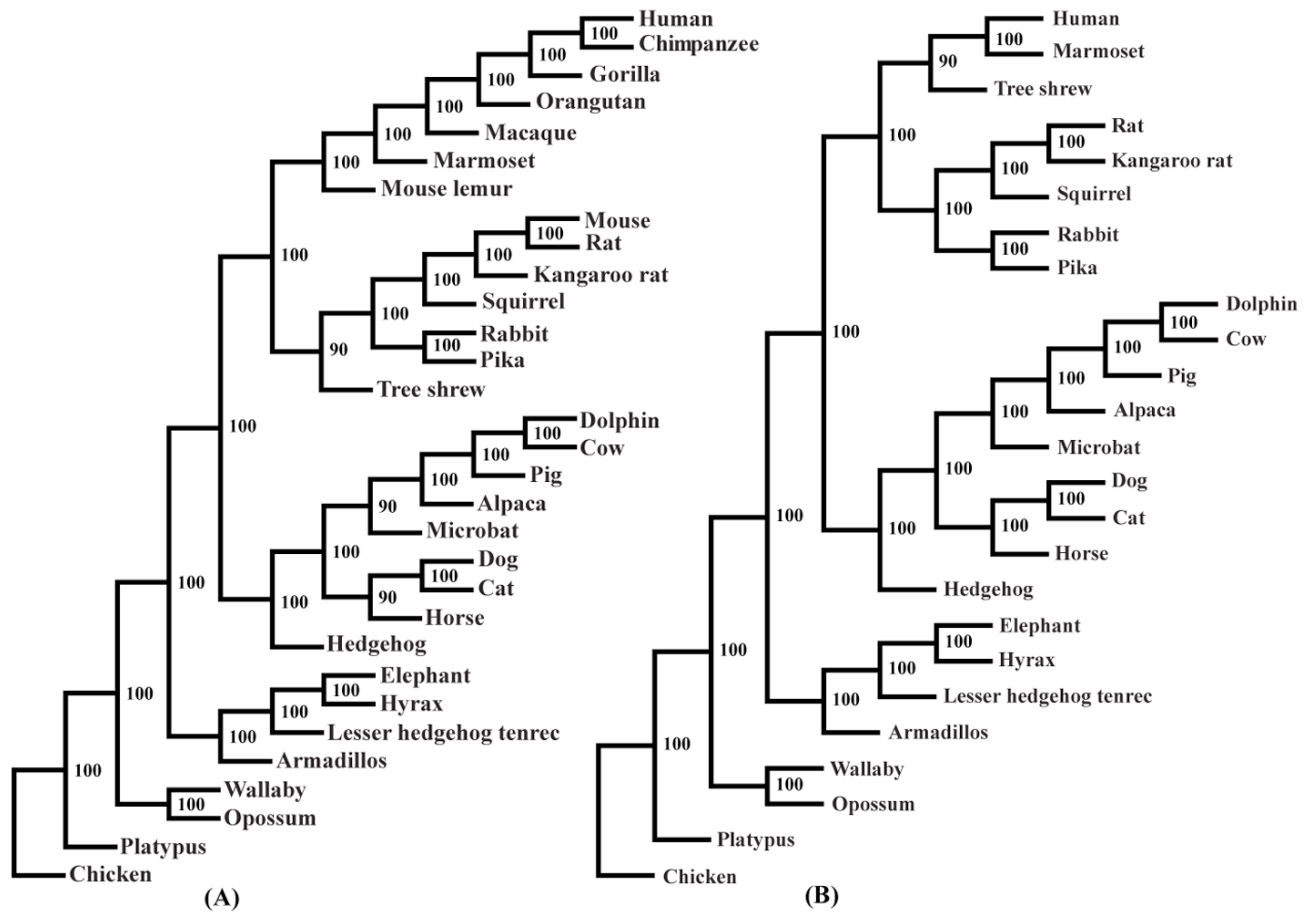
Supplementary Fig. S5. Eutherian phylogeny estimated with reduced taxon sampling using the coalescent method MP-EST. A. Tree with six taxa excluded; B. Tree with twelve taxa excluded. The numbers on the tree indicate the bootstrap support values for each of the corresponding nodes.



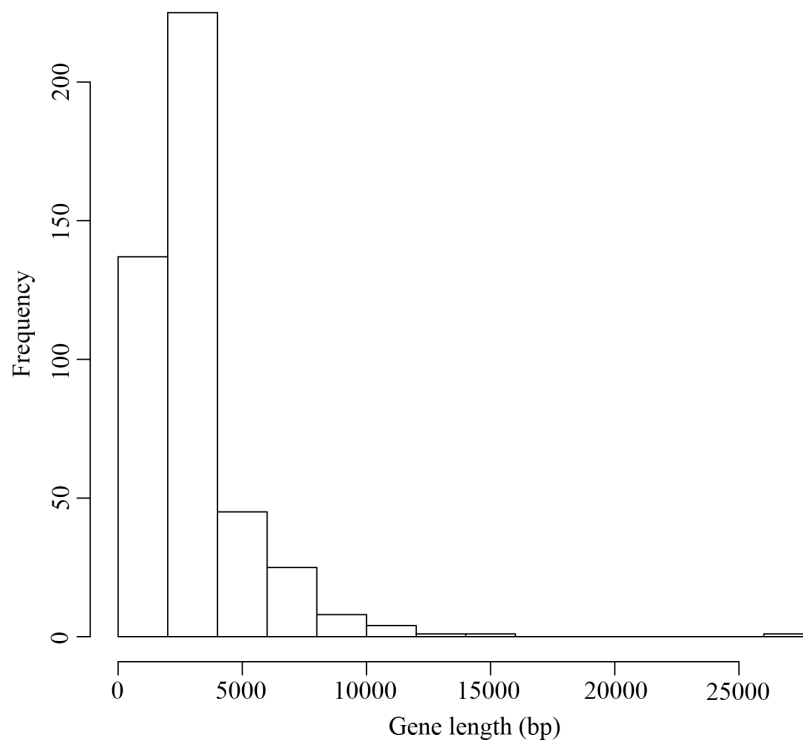
Supplementary Fig. S6. Eutherian phylogeny estimated with reduced taxon sampling using the coalescent method STAR. A. Tree with six taxa excluded; B. Tree with twelve taxa excluded. The numbers on the tree indicate the bootstrap support values for each of the corresponding nodes.



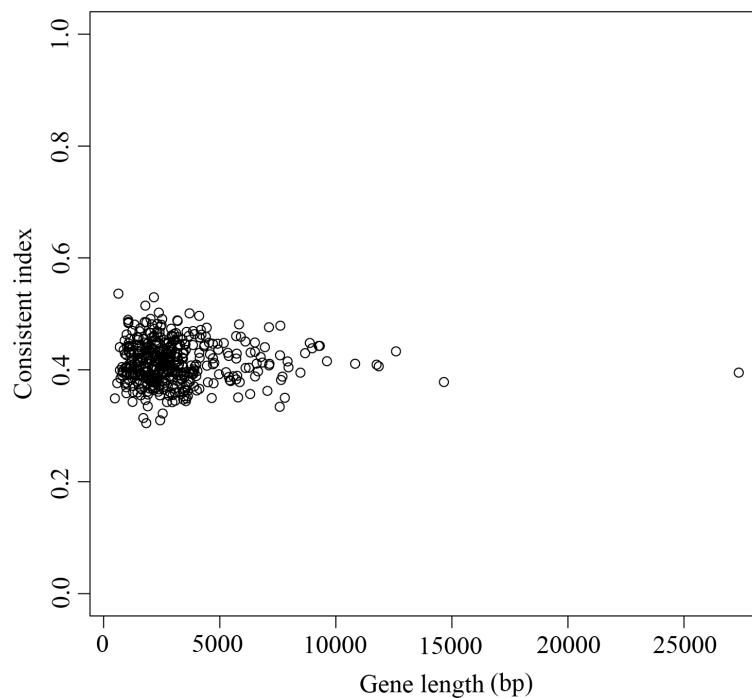
Supplementary Fig. S7. Eutherian phylogeny estimated with reduced taxon sampling using the concatenation method MrBayes. A. Tree with six taxa excluded; B. Tree with twelve taxa excluded. The numbers on the tree indicate posterior probability percentages for each of the corresponding nodes.



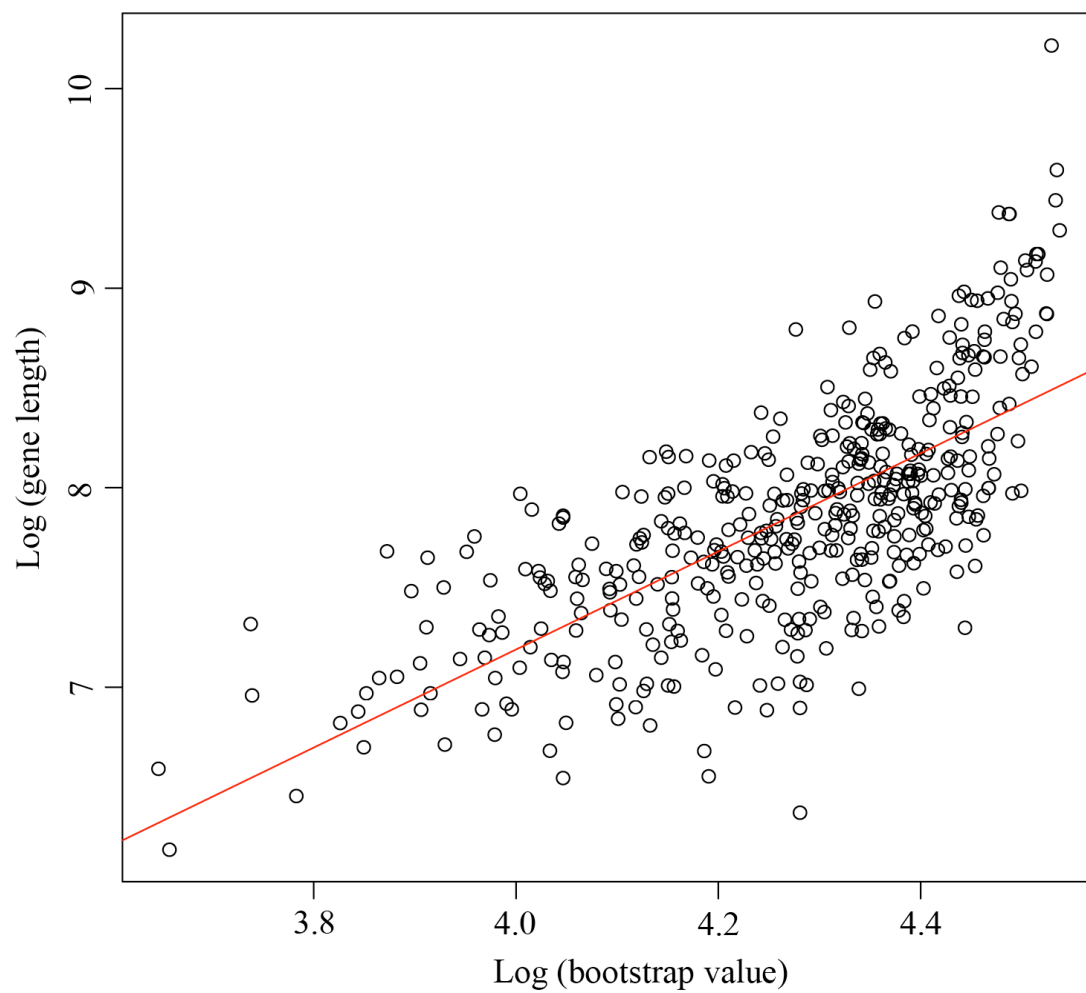
Supplementary Fig. S8. Eutherian phylogeny estimated with reduced taxon sampling using the ML concatenation method RAXML. A. Tree with six taxa excluded; B. Tree with twelve taxa excluded. The numbers on the tree indicate bootstrap support values for each of the corresponding nodes.



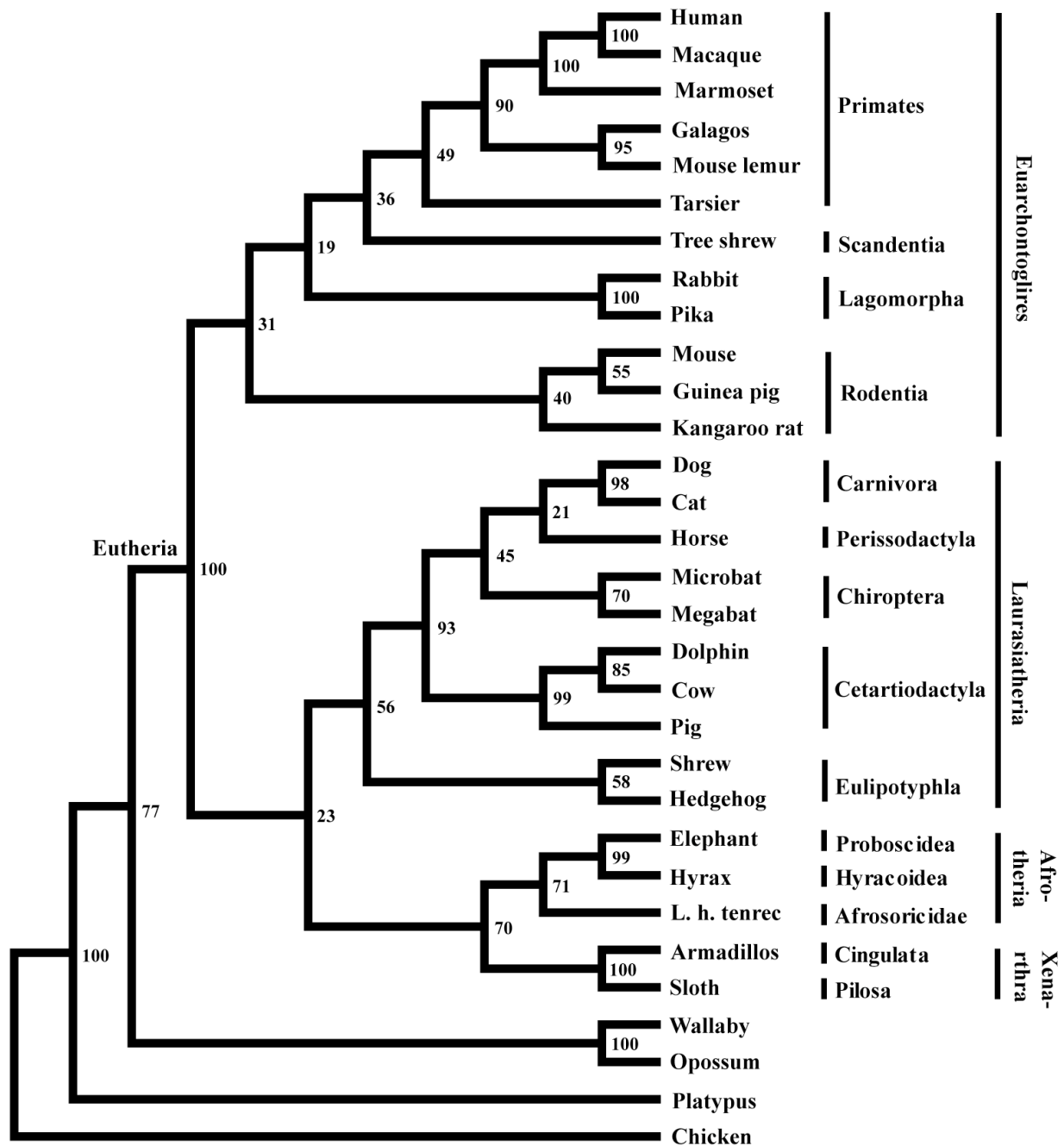
Supplementary Fig. S9. The length of genes sampled for this study.



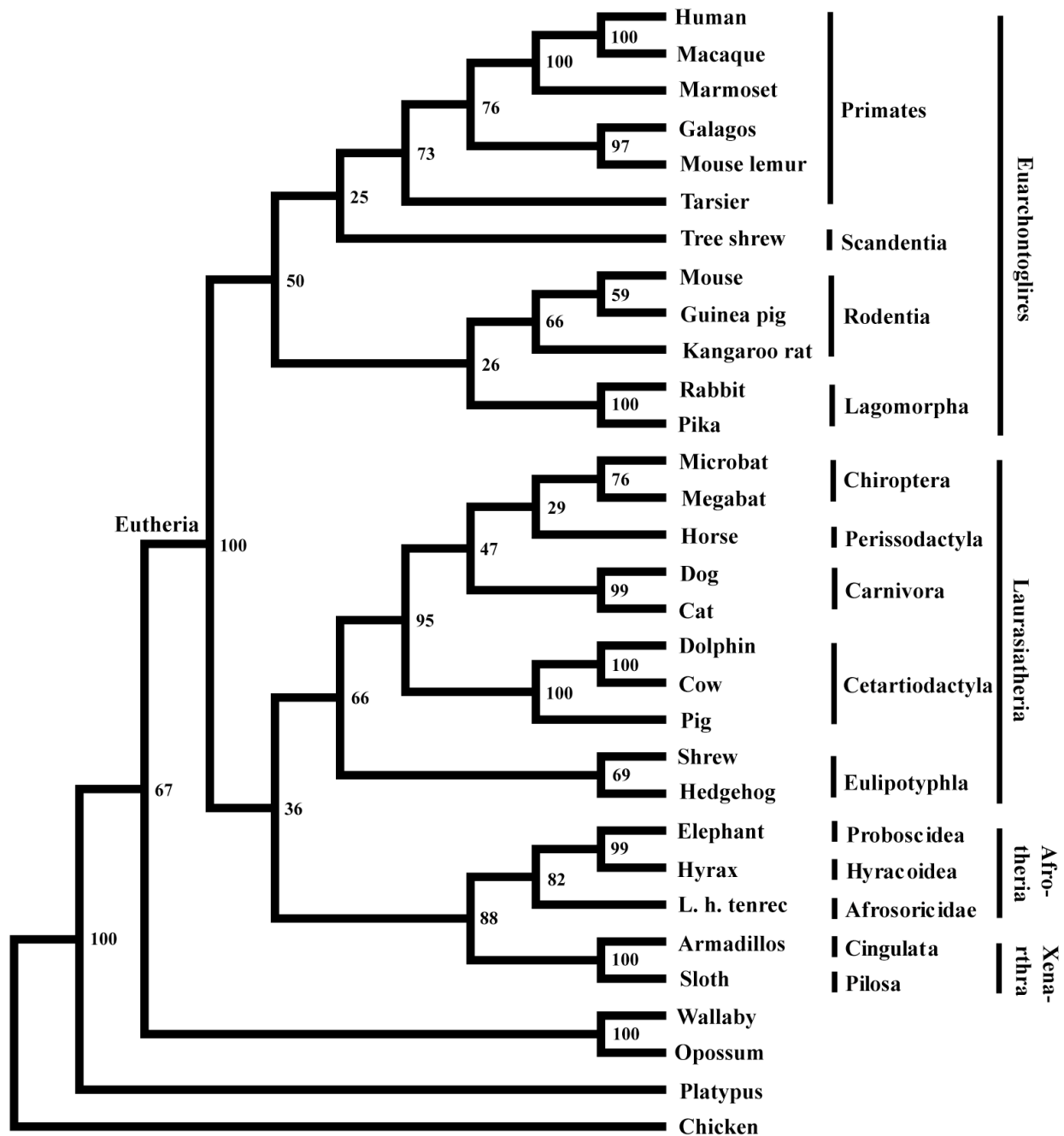
Supplementary Fig. S10. Correlation between gene length and consistency index among the 447 loci. There is no correlation between gene length and consistency index (p -value = 0.686), indicating that the data set contains no systematic effect of recombination.



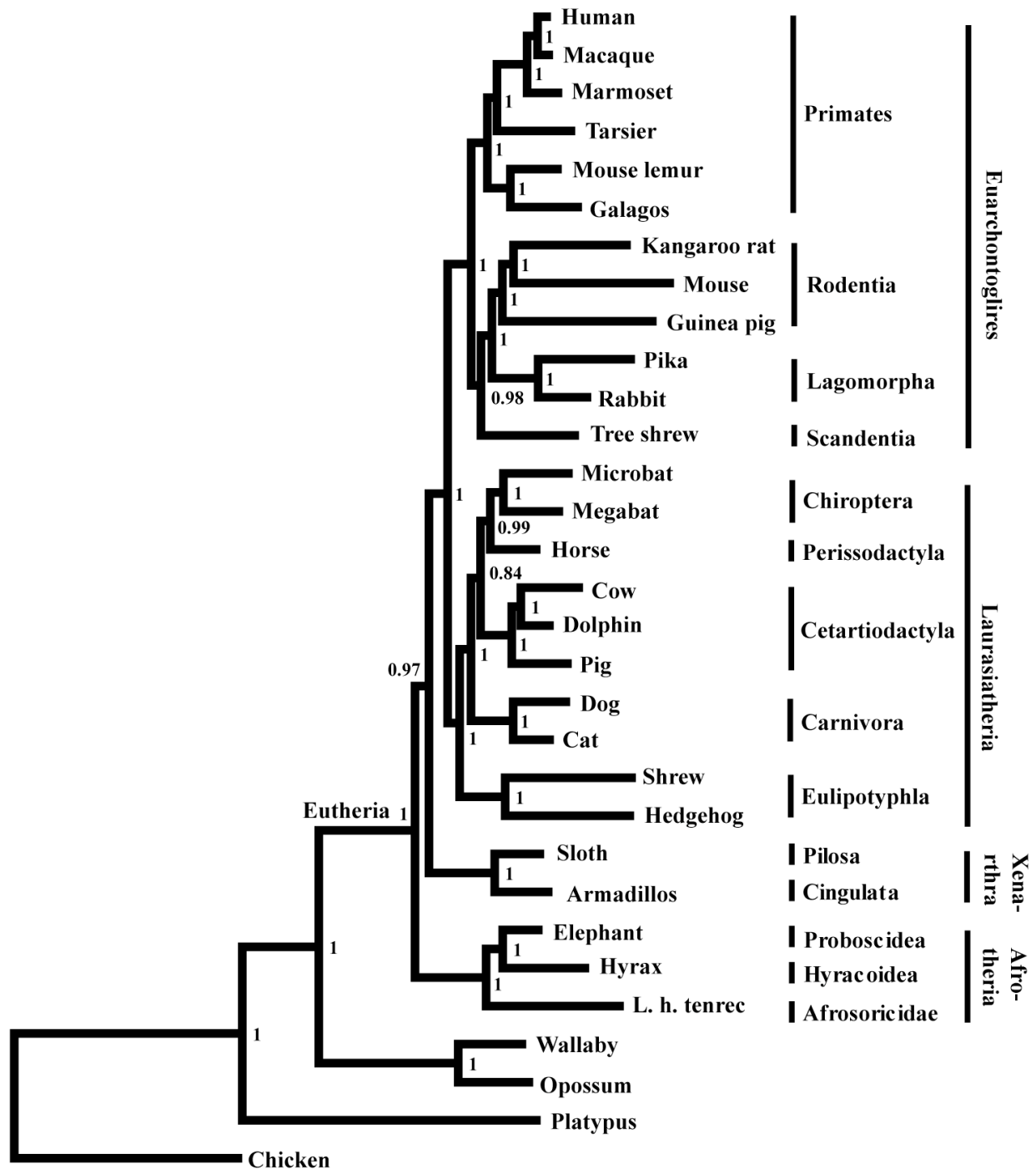
Supplementary Fig. S11. Correlation between gene length and average bootstrap value for each of the 447 gene trees. There is a strong correlation between gene length and average bootstrap value (p-value <0.001).



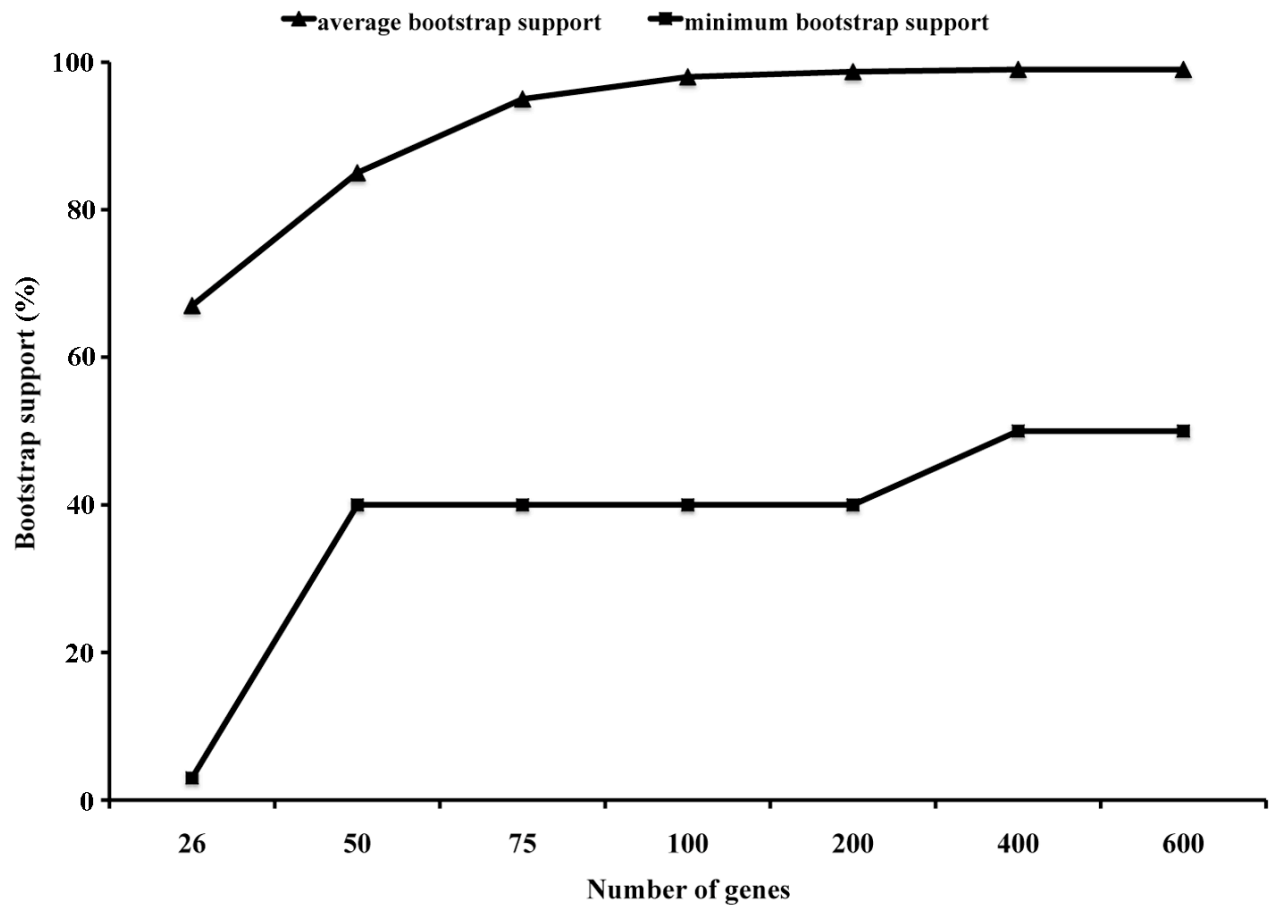
Supplementary Fig. S12. Eutherian phylogeny estimated using the coalescent method MP-EST based on the genetic data set (26 genes) of Meredith et al.(12) and the 31 taxa shared between this study and Meredith et al. The numbers on the tree indicate the bootstrap support values for each of the nodes. Abbreviation: L. h. tenrec, lesser hedgehog tenrec.



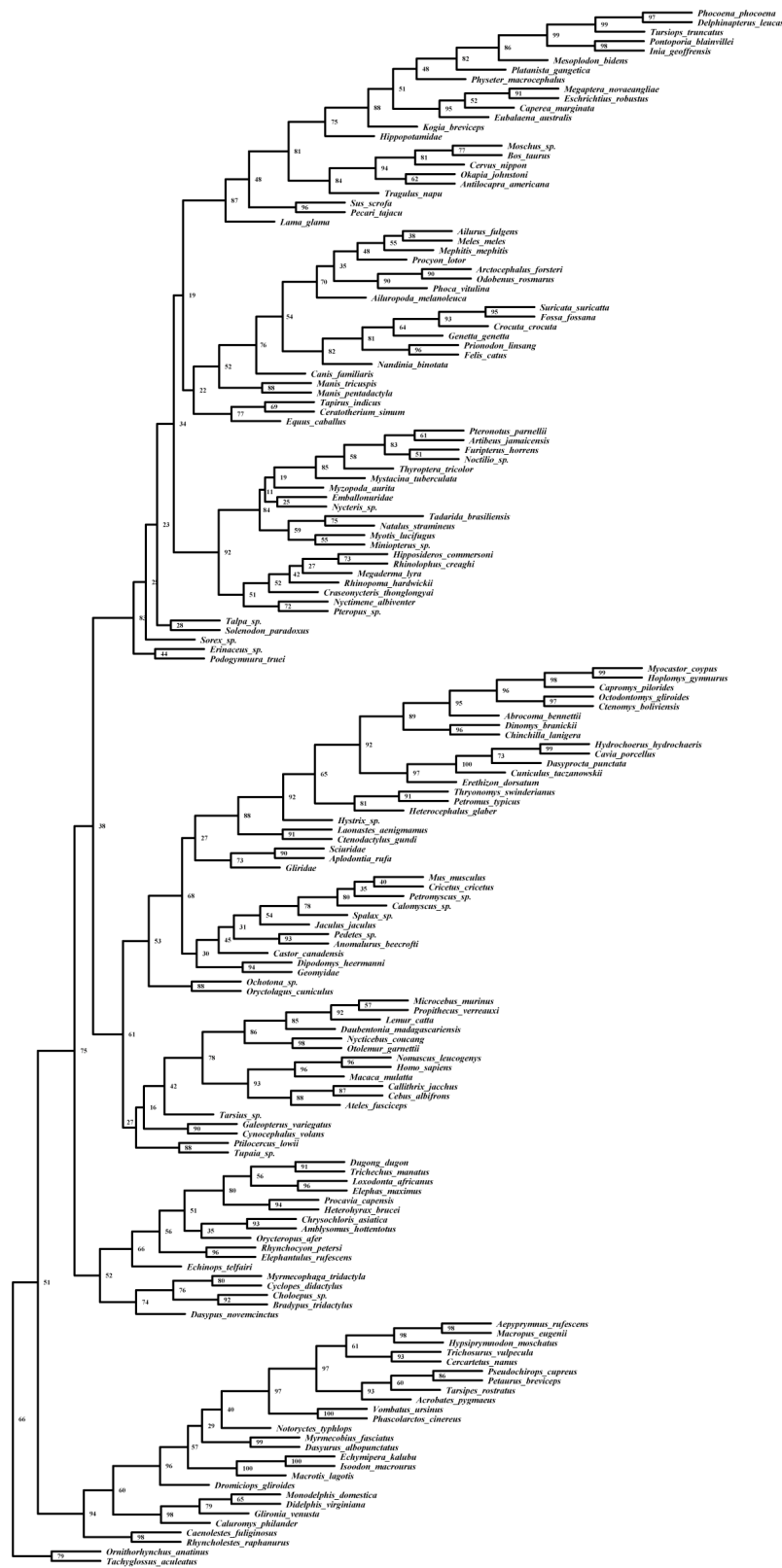
Supplementary Fig. S13. Eutherian phylogeny estimated using the coalescent method STAR based on the genetic data set (26 genes) of Meredith et al.(12) and the 31 taxa shared between this study and Meredith et al. The numbers on the tree indicate the bootstrap support values for each of the nodes.



Supplementary Fig. S14. Eutherian phylogeny estimated using the concatenation method MrBayes based on the genetic data set (26 genes) of Meredith et al.(12) and 31 taxa shared between this study and Meredith et al. The numbers on the tree indicate the posterior probability percentages for each of the nodes.



Supplementary Fig. S15. Estimation of the number of genes for resolving the eutherian tree (169 taxa) of Meredith et al. 2011 with high confidence. The results suggest that it would require at least 400 genes to resolve the eutherian tree with an average bootstrap support of >95% and a minimum bootstrap support of to 50%.



Supplementary Fig. S16. Eutherian tree estimated using the coalescent method MP-EST based on the full data set of Meredith et al. (12). Non-mammalian taxa were not shown. The numbers on the tree indicate the bootstrap support values.

Supplementary Tables

Supplementary Table S1. List of taxon sampling for this study

Order	Family	Species	Common Name
Artiodactyla	Camelidae	<i>Vicugna pacos</i>	Alpaca
Artiodactyla	Bovidae	<i>Bos taurus</i>	Cow
Artiodactyla	Suidae	<i>Sus scrofa</i>	Pig
Perissodactyla	Equidae	<i>Equus caballus</i>	Horse
Carnivora	Canidae	<i>Canis familiaris</i>	Dog
Carnivora	Felidae	<i>Felis catus</i>	Cat
Cetacea	Delphinidae	<i>Tursiops truncatus</i>	Dolphin
Erinaceomorpha	Erinaceidae	<i>Erinaceus europaeus</i>	Hedgehog
Chiroptera	Pteropodidae	<i>Pteropus vampyrus</i>	Megabat
Chiroptera	Vespertilionidae	<i>Myotis lucifugus</i>	Microbat
Soricomorpha	Soricidae	<i>Sorex araneus</i>	Shrew
Scandentia	Tupaiaidae	<i>Tupaia belangeri</i>	Tree shrew
Primates	Galagonidae	<i>Otolemur garnettii</i>	Galagos
Primates	Hominidae	<i>Pan troglodytes</i>	Chimpanzee
Primates	Hominidae	<i>Gorilla gorilla</i>	Gorilla
Primates	Hominidae	<i>Homo sapiens</i>	Human
Primates	Cercopithecidae	<i>Macaca mulatta</i>	Macaque
Primates	Callitrichidae	<i>Callithrix jacchus</i>	Marmoset
Primates	Cheirogaleidae	<i>Microcebus murinus</i>	Mouse Lemur
Primates	Hominidae	<i>Pongo pygmaeus</i>	Orangutan
Primates	Tarsiidae	<i>Tarsius syrichta</i>	Tarsier
Rodentia	Hyteromyidae	<i>Dipodomys ordii</i>	Kangaroo rat
Rodentia	Muridae	<i>Mus musculus</i>	Mouse
Rodentia	Muridae	<i>Rattus</i>	Rat
Rodentia	Sciuridae	<i>Spermophilus tridecemlineatus</i>	Squirrel
Rodentia	Caviidae	<i>Cavia porcellus</i>	Guinea Pig
Lagomorpha	Ochotonidae	<i>Ochotona princeps</i>	Pika
Lagomorpha	Leporidae	<i>Oryctolagus cuniculus</i>	Rabbit
Proboscidea	Elephantidae	<i>Loxodonta africana</i>	Elephant
Hyracoidea	Procaviidae	<i>Procavia capensis</i>	Hyrax
Afrosoricidae	Tenrecidae	<i>Echinops telfairi</i>	Lesser Hedgehog Tenrec
Cingulata	Dasypodidae	<i>Dasypus novemcinctus</i>	Armadillos
Pilosa	Megalonychidae	<i>Choloepus hoffmanni</i>	Sloth
Didelphimorphia	Didelphidae	<i>Monodelphis domestica</i>	Opossum
Diprotodontia	Macropodidae	<i>Macropus eugenii</i>	Wallaby
Monotremata	Ornithorhynchidae	<i>Ornithorhynchus anatinus</i>	Platypus
		<i>Gallus gallus</i>	Chicken

Supplementary Table S2. List of the 447 nuclear loci sampled and their distribution in the human genome

Gene	Position	Length (bp)	Gene	Position	Length (bp)
ZEB1	10p11.2	3420	MEI1	22q13.2	4080
CREM	10p11.21	1020	BRPF1	22q13.33	3960
PARD3	10p11.21	4080	BRPF1	22q13.33	3960
NRP1	10p12	2820	PTCD3	2p11.2	2160
NEBL	10p12	3240	EIF2AK3	2p12	3420
CACNB2	10p12	1860	MXD1	2p13-p12	720
ARMC3	10p12.31	2760	ASB3	2p16-p14	1680
PLXDC2	10p12.31	1560	MTIF2	2p16.1	2160
DCLRE1C	10p13	2400	EML4	2p21	3060
ITGA8	10p13	3240	SRBD1	2p21	3180
STAM	10p14-p13	1680	CCDC75	2p22.2	960
RBM17	10p15.1	1260	HADHA	2p23	2280
PRKG1	10q11.2	2040	AGBL5	2p23.3	2940
ERCC6	10q11.23	5400	DDX1	2p24	2340
BICC1	10q21.1	2760	ASXL2	2p24.1	4680
RHOBTB1	10q21.2	2100	MYO1B	2q12-q34	3420
SLC25A16	10q21.3	1020	POLR1B	2q13	3420
BTA1F1	10q22-q23	5700	MERTK	2q14.1	3300
SUPV3L1	10q22.1	2460	R3HDM1	2q21.3	3660
KIAA1279	10q22.1	1860	NR4A2	2q22-q23	1740
MYST4	10q22.2	6300	SLC4A10	2q23-q24	3360
KCNMA1	10q22.3	3720	IFIH1	2q24	3180
PLCE1	10q23	7140	LASS6	2q24.3	1140
DNTT	10q23-q24	1620	COL3A1	2q31	4680
SORCS1	10q23-q25	3480	DYNC112	2q31.1	1980
IDE	10q23-q25	3180	GPR155	2q31.1	2700
SH2D4B	10q23.1	1320	DYNC112	2q31.1	1980
PDE6C	10q24	2580	GORASP2	2q31.1-q31.2	1380
GBF1	10q24	5700	GLS	2q32-q34	2160
MYOF	10q24	6600	CALCRL	2q32.1	1440
ENTPD1	10q24	1620	ABCA12	2q34	8940
NRAP	10q24-q26	5220	COL4A4	2q35-q37	5700
KIF11	10q24.1	3240	SPHKAP	2q36	5460
MGEA5	10q24.1-q24.3	2820	FARSB	2q36.1	1800
SHOC2	10q25	1740	RHBDD1	2q36.3	960
TECTB	10q25-q26	960	PSMD1	2q37.1	2940
GPAM	10q25.2	2520	GLT8D4	3p13	1080
RBM20	10q25.2	2700	C3orf67	3p14.2	2580
C10orf134	10q25.3	1920	PBRM1	3p21	5160
HABP2	10q25.3	1800	PTPRG	3p21-p14	4440
HSPA12A	10q26.12	2040	DOCK3	3p21.2	6240
CUGBP1	11p11	1440	RBM6	3p21.3	3660
NDUFS3	11p11.11	780	EFHB	3p24.3	2100

Supplementary Table S2. (Continued 1)

KIAA0652	11p11.2	1680	ANKRD28	3p25.1	3480
FBNP4	11p11.2	3300	C3orf15	3q12-q13.3	2400
CD44	11p13	2580	TOMM70A	3q12.2	1860
KIF18A	11p14.1	2760	CASR	3q13	3360
SLC5A12	11p14.2	1920	ALCAM	3q13.1	1800
PAN3	11p15	2220	ADPRH	3q13.31-q13.33	1140
NAV2	11p15.1	7680	ACPP	3q21-q23	1320
ABCC8	11p15.1	4920	NEK11	3q22.1	2160
COPB1	11p15.2	2820	FND3B	3q26.31	3660
INSC	11p15.2	1740	PEX5L	3q26.33	1860
SBF2	11p15.4	5700	TTC14	3q26.33	2400
PPFIBP2	11p15.4	3120	YEATS2	3q27.1	4560
SBF2	11p15.4	5700	MTF1	3q28-q29	2340
FCHSD2	11q13.4	2160	ATP13A3	3q29	3840
C11orf30	11q13.5	4080	LSG1	3q29	2160
ANKRD42	11q14.1	1560	APBB2	4p13	2340
C11orf82	11q14.1	3180	STIM2	4p15.2	2160
CHORDC1	11q14.3	960	SLIT2	4p15.2	4620
MTMR2	11q22	1980	PACRGL	4p15.31	780
ATM	11q22-q23	9300	PDGFRA	4q12	3300
FDXACB1	11q23.1	1920	MPDZ	4q12	6480
ZW10	11q23.2	2400	UBA6	4q13.2	3180
RNF214	11q23.3	2160	RASSF6	4q13.3	1080
FEZ1	11q24.2	1260	AFF1	4q21	4320
DNM1L	12p11.21	2280	SLC4A4	4q21	3300
SOX5	12p12.1	2280	SCARB2	4q21.1	1440
LPCAT3	12p13	1500	HELQ	4q21.23	3360
ATF7IP	12p13.1	4080	KLHL8	4q22.1	1860
ERC1	12p13.3	3360	MAPK10	4q22.1-q23	1500
LRIG3	12q14.1	3600	CFI	4q25	2100
EPYC	12q21	960	TRPC3	4q27	2880
E2F7	12q21.2	2880	HHIP	4q28-q32	2100
NR2C1	12q22	1920	PRMT10	4q31.23	2580
APAF1	12q23	3900	PET112L	4q31.3	1680
C12orf63	12q23.1	3840	GUCY1B3	4q31.3-q33	1860
SLC17A8	12q23.1	1800	TMEM144	4q32.1	1080
AMDHD1	12q23.1	1320	CLCN3	4q33	2640
STAB2	12q23.3	7920	GPM6A	4q34	960
SART3	12q24.1	3060	GALNTL6	4q34.1	1800
KNTC1	12q24.31	6840	ENPP6	4q35.1	1380
SFRS8	12q24.33	2940	CCDC111	4q35.1	1860
ALOX5AP	13q12	480	SNX25	4q35.1	2700
FLT3	13q12	3180	NNT	5p12	3360
FLT3	13q12	3180	AGXT2	5p13	1620
LNX2	13q12.2	2520	IL7R	5p13	1440

Supplementary Table S2. (Continued 2)

B3GALTL	13q12.3	1440	LIFR	5p13-p12	3420
SPG20	13q13.3	2220	FYB	5p13.1	2640
POSTN	13q13.3	2580	C5orf51	5p13.1	900
LRCH1	13q14.11	2520	PRLR	5p13.2	2520
CPB2	13q14.11	1260	EGFLAM	5p13.2-p13.1	3180
PIBF1	13q22.1	2340	MTMR12	5p13.3	2280
LMO7	13q22.2	6060	CTNND2	5p15.2	3780
CLYBL	13q32	1140	NSUN2	5p15.31	2700
NUBPL	14q12	960	IL6ST	5q11.2	2880
DDHD1	14q21	3000	MAST4	5q12.3	7920
PYGL	14q21-q22	2580	ERBB2IP	5q12.3	4380
SEC23A	14q21.1	2280	PIK3R1	5q13.1	2220
TRIM9	14q22.1	2580	AP3B1	5q14.1	3480
C14orf101	14q22.3	2160	HOMER1	5q14.2	1140
GPHN	14q23.3	2460	VCAN	5q14.3	12540
SIPA1L1	14q24.2	5760	RHOBTB3	5q15	1860
YLPM1	14q24.3	7560	APC	5q21-q22	8640
SEL1L	14q31	2400	MAN2A1	5q21-q22	3720
TDP1	14q32.11	1920	DMXL1	5q22	9240
TRAF3	14q32.32	1680	YTHDC2	5q22.2	4380
GABRB3	15q11.2-q12	1560	TRIM36	5q22.3	2280
AVEN	15q13.1	1080	DTWD2	5q23.1	900
AQR	15q14	4680	GRAMD3	5q23.2	1380
PDIA3	15q15	1500	MYOT	5q31	2160
TLN2	15q15-q21	7800	CDH20	5q31	2460
GANC	15q15.2	2820	ACSL6	5q31	2280
CASC4	15q15.3	1320	ARHGAP26	5q31	2460
RNF111	15q21	3060	ITK	5q31-q32	1980
UNC13C	15q21.3	6540	IL12B	5q31.1-q33.1	1080
DENND4A	15q22.31	5880	DPYSL3	5q32	2040
TM6SF1	15q24-q26	1140	MEGF10	5q33	3480
SIN3A	15q24.2	3900	CLINT1	5q33.3	1980
FANCI	15q26.1	4080	TINAG	6p12.1	1500
MCTP2	15q26.2	2760	UBR2	6p21.1	5340
ARHGAP17	16p12.1	2760	SUPT3H	6p21.1-p21.3	1200
ARHGAP17	16p12.1	2760	FTSJD2	6p21.2	2580
EEF2K	16p12.2	2160	C6orf89	6p21.2	1080
TXNDC11	16p13.13	2880	DNAH8	6p21.2	14640
CREBBP	16p13.3	7620	AOF1	6p22.3	2580
ITFG1	16q12.1	1860	GMPR	6p23	1200
RPGRIP1L	16q12.2	4140	EDN1	6p24.1	660
GIN53	16q21	780	C6orf105	6p24.1	900
EDC4	16q22.1	4860	SLC35B3	6p24.3	1200
TAT	16q22.1	1320	NEDD9	6p25-p24	2940
PRMT7	16q22.1	2100	NEDD9	6p25-p24	2940

Supplementary Table S2. (Continued 3)

SF3B3	16q22.1	3720	RIPK1	6p25.2	2220
TERF2	16q22.1	2280	ZNF318	6pter-p12.1	7080
PHLPP2	16q22.2	4020	PHF3	6q12	6480
GLG1	16q22.3	3840	COL12A1	6q12-q13	9600
VAT1L	16q23.1	1260	COL12A1	6q12-q13	9600
WDR59	16q23.1	3060	LCA5	6q14.1	2280
GAN	16q24.1	2820	SNAP91	6q14.2	2880
GLP2R	17p13.3	1680	AKIRIN2	6q15	600
TSR1	17p13.3	2880	SIM1	6q16.3-q21	2460
SMYD4	17p13.3	2520	CDC2L6	6q21	1560
PRPF8	17p13.3	7020	MYB	6q22-q23	2400
KIAA0100	17q11.2	6900	ENPP1	6q22-q23	2880
BLMH	17q11.2	1380	LACE1	6q22.1	1500
TAOK1	17q11.2	3060	C6orf192	6q22.3-q23.3	1440
SSH2	17q11.2	4500	L3MBTL3	6q23	3540
TMIGD1	17q11.2	780	EYA4	6q23	1980
AATF	17q12	1860	HBS1L	6q23.3	2100
APPBP2	17q21-q23	1800	GPR126	6q24.1	3900
TTC25	17q21.2	2820	TXLNB	6q24.1	3180
ANKFN1	17q22	2280	SYNE1	6q25	27300
TUBD1	17q23.1	1440	C6orf97	6q25.1	2100
PPM1D	17q23.2	2040	SERAC1	6q25.3	1980
PITPNC1	17q24.2	960	QKI	6q26	1020
NUP85	17q25.1	2040	AHR	7p15	2760
SPIRE1	18p11.21	2100	OSBPL3	7p15	2820
LPIN2	18p11.31	3000	RAPGEF5	7p15.3	2520
ANKRD29	18q11.2	900	ANKMY2	7p21	1440
NOL4	18q12	1980	THSD7A	7p21.3	5340
FHOD3	18q12	4740	SCIN	7p21.3	2400
MOCOS	18q12	2700	ECD	7q21.3	1980
FAM59A	18q12.1	2580	LAMB1	7q22	5700
ATP8B1	18q21-q22	3840	SMURF1	7q22.1	2400
SETBP1	18q21.1	4680	NRCAM	7q31	4440
RLF	19p13.2-p12	6060	COG5	7q31	2580
GPATCH1	19q13.11	2880	SLC13A1	7q31-q32	1980
ZNF507	19q13.11	2880	CADPS2	7q31.3	3960
WDR3	1p12	2880	TNPO3	7q32.1	3060
MAN1A2	1p13	2040	DGKI	7q32.3-q33	3060
RSBN1	1p13.2	2520	WDR91	7q33	2340
VAV3	1p13.3	2520	SLC35B4	7q33	960
SLC25A24	1p13.3	1440	NCAPG2	7q36.3	3480
ALG14	1p21.3	660	LETM2	8p11.23	1440
ABCA4	1p22	7080	UNC5D	8p12	2940
CLCA2	1p22.3	2880	ENTPD4	8p21.3	1860
DLSTP	1p31	1440	SLC18A1	8p21.3	1560
PDE4B	1p31	2220	ZDHHC2	8p22	1140

Supplementary Table S2. (Continued 4)

IL12RB2	1p31.3-p31.2	2940	MTUS1	8p22	4200
C8A	1p32	1860	DRP2	8p22-p21	3120
EPS15	1p32	2940	TNKS	8p23.1	3960
ORC1L	1p32	2880	MCM4	8q11.2	2640
CSMD2	1p34.3	11820	NSMAF	8q12-q13	2760
EYA3	1p36	1740	GGH	8q12.3	960
GRHL3	1p36.11	2220	CYP7B1	8q21.3	1560
KIAA0090	1p36.13	3000	DPYS	8q22	1560
CASP9	1p36.21	1680	MTERFD1	8q22.1	1260
ACBD6	1q25.1	840	SLC26A7	8q23	2040
RALGPS2	1q25.2	1740	ADCY8	8q24	3840
DSTYK	1q32.1	2880	ANXA13	8q24.13	1080
IARS2	1q41	3180	KIAA0196	8q24.13	3540
TP53BP2	1q41	3540	LRRC6	8q24.22	1440
IQCH	1q41	3420	PTK2	8q24.3	3600
QARS	1q41	2340	MELK	9p13.2	2040
C1orf58	1q41	1200	GNE	9p13.3	2160
TAF1A	1q42	1440	TEK	9p21	3480
LYST	1q42.1-q42.2	11700	PLAA	9p21	2400
LYST	1q42.1-q42.2	11700	TTC39B	9p22.3	1860
CDC42BPA	1q42.11	5340	FREM1	9p22.3	6720
C1orf55	1q42.12	1500	TYRP1	9p23	1680
SDCCAG8	1q43	2340	SLC1A1	9p24	1560
XRN2	20p11.2-p11.1	2880	ERMP1	9p24	2880
C20orf74	20p11.22	6480	SMC5	9q21.12	3360
C20orf72	20p11.23	1080	TRPM3	9q21.12	5580
CSRP2BP	20p11.23	2400	TMC1	9q21.12	2640
PLCB4	20p12	3600	PRUNE2	9q21.2	8820
PAK7	20p12	2160	RNF20	9q22	2940
MCM8	20p12.3	2760	TTF2	9q22	3900
ATRN	20p13	4320	C9orf5	9q31	2880
DHX35	20q11.22-q12	2160	GRIN3A	9q31.1	3480
CHD6	20q12	8460	SVEP1	9q32	10800
TTPAL	20q13.12	1080	FKBP15	9q32	4140
ADNP	20q13.13	3360	PAPPA	9q33.2	4920
ARFGEF2	20q13.13	5820	GOLGA1	9q33.3	2340
RAB22A	20q13.32	540	SPTAN1	9q34.11	7560
HSPA13	21q11.1	1440	OTC	Xp21.1	1020
ADAMTS5	21q21.3	2880	CDKL5	Xp22	3420
C21orf66	21q21.3	2820	PIGA	Xp22.1	1500
KCNJ6	21q22.1	1200	IL1RAPL1	Xp22.1-p21.3	2220
MPP7	21q22.13	1800	CENPI	Xq22.1	2400
DSCAM	21q22.2	6300	PLS3	Xq23	1920
TNRC6B	22q13.1	5820	TRPC5	Xq23	3000
EP300	22q13.2	7560			

Supplementary Table S3. List of eutherian taxa excluded for the phylogenetic analyses with reduced taxon sampling.

Six taxa excluded		Twelve taxa excluded	
Species	Common Name	Species	Common Name
<i>Otolemur garnettii</i>	Galagos	<i>Otolemur garnettii</i>	Galagos
<i>Tarsius syrichta</i>	Tarsier	<i>Tarsius syrichta</i>	Tarsier
<i>Cavia porcellus</i>	Guinea Pig	<i>Pan troglodytes</i>	Chimpanzee
<i>Sorex araneus</i>	Shrew	<i>Gorilla gorilla</i>	Gorilla
<i>Pteropus vampyrus</i>	Megabat	<i>Pongo pygmaeus</i>	Orangutan
<i>Choloepus hoffmanni</i>	Sloth	<i>Macaca mulatta</i>	Macaque
		<i>Microcebus murinus</i>	Mouse Lemur
		<i>Mus musculus</i>	Mouse
		<i>Cavia porcellus</i>	Guinea Pig
		<i>Sorex araneus</i>	Shrew
		<i>Pteropus vampyrus</i>	Megabat
		<i>Choloepus hoffmanni</i>	Sloth

Supplementary Table S4. The 31 taxa shared between this study and Meredith et al. 2011(12).

Order	Family	Species	Common Name
Artiodactyla	Bovidae	<i>Bos taurus</i>	Cow
Artiodactyla	Suidae	<i>Sus scrofa</i>	Pig
Perissodactyla	Equidae	<i>Equus caballus</i>	Horse
Carnivora	Canidae	<i>Canis familiaris</i>	Dog
Carnivora	Felidae	<i>Felis catus</i>	Cat
Cetacea	Delphinidae	<i>Tursiops truncatus</i>	Dolphin
Erinaceomorpha	Erinaceidae	<i>Erinaceus europaeus</i>	Hedgehog
Chiroptera	Pteropodidae	<i>Pteropus vampyrus</i>	Megabat
Chiroptera	Vespertilionidae	<i>Myotis lucifugus</i>	Microbat
Soricomorpha	Soricidae	<i>Sorex araneus</i>	Shrew
Scandentia	Tupauidae	<i>Tupaia belangeri</i>	Tree shrew
Primates	Galagonidae	<i>Otolemur garnettii</i>	Galagos
Primates	Hominidae	<i>Homo sapiens</i>	Human
Primates	Cercopithecidae	<i>Macaca mulatta</i>	Macaque
Primates	Callitrichidae	<i>Callithrix jacchus</i>	Marmoset
Primates	Cheirogaleidae	<i>Microcebus murinus</i>	Mouse Lemur
Primates	Tarsiidae	<i>Tarsius syrichta</i>	Tarsier
Rodentia	Hyteromyidae	<i>Dipodomys ordii</i>	Kangaroo rat
Rodentia	Muridae	<i>Mus musculus</i>	Mouse
Rodentia	Caviidae	<i>Cavia porcellus</i>	Guinea Pig
Lagomorpha	Ochotonidae	<i>Ochotona princeps</i>	Pika
Lagomorpha	Leporidae	<i>Oryctolagus cuniculus</i>	Rabbit
Proboscidea	Elephantidae	<i>Loxodonta africana</i>	Elephant
Hyracoidea	Procaviidae	<i>Procavia capensis</i>	Hyrax
Afrosoricidae	Tenrecidae	<i>Echinops telfairi</i>	lesser hedgehog tenrec
Cingulata	Dasypodidae	<i>Dasypus novemcinctus</i>	Armadillos
Pilosa	Megalonychidae	<i>Choloepus hoffmanni</i>	Sloth
Didelphimorphia	Didelphidae	<i>Monodelphis domestica</i>	Opossum
Diprotodontia	Macropodidae	<i>Macropus eugenii</i>	Wallaby
Monotremata	Ornithorhynchidae	<i>Ornithorhynchus anatinus</i>	Platypus
		<i>Gallus gallus</i>	Chicken

Supplementary Table S5. Comparisons of the average bootstrap values for the eutherian tree estimated using MP-EST with the longest and shortest 50, 100 and 200 gene sets from the 447 genes.

Gene set	Average bootstrap value (%)	
	Longest	Shortest
50	90.71	81.08
100	93.26	90.63
200	94.43	92.54

References

1. Ranwez V, *et al.* (2007) OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* 7(241):doi: 10.1186/1471-2148-1187-1241.
2. Knight R, *et al.* (2007) PyCogent: a toolkit for making sense from sequence. *Genome Biol.* 8:R171.
3. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-1797.
4. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.
5. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690.
6. Robinson DF & Foulds LR (1981) Comparison of phylogenetic trees. *Math. Biosci.* 53:131-147.
7. Margush T & McMorris FR (1981) Consensus n-trees. *Bull. Math. Biol.* 43:239-244.
8. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164-166.
9. Ané C (2010) Reconstructing concordance trees and testing the coalescent model from genome-wide data sets. *Estimating Species Trees: Practical and Theoretical Aspects*, eds Knowles LL & Kubatko L (Wiley-Blackwell, New Jersey), pp 35-52.
10. Liu L, Yu L, & Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* 10:302.
11. Kluge AG & Farris JS (1969) Quantitative phyletics and the evolution of anurans. *Syst. Biol.* 18:1-32.
12. Meredith RW, *et al.* (2011) Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* 334:521-524.